Development 112, 335–350 (1991)
Printed in Great Britain © The Company of Biologists Limited 1991

Gene regulatory factors of the sea urchin embryo

I. Purification by affinity chromatography and cloning of P3A2, a novel DNA-binding protein

FRANK J. CALZONE*, CHRISTER HÖÖG†, DAVID B. TEPLOW, ANN E. CUTTING, ROBERT W. ZELLER, ROY J. BRITTEN and ERIC H. DAVIDSON

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

* Present address Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92717, USA
† Present address Department of Molecular Genetics, Karolinska Institutet, S-10401 Stockholm, Sweden

Summary

The P3A2 regulatory protein interacts with specific sites in the control region of the *CyIIIa* actin gene. Previous studies showed that this interaction is required to confine expression of a *CyIIIa · CAT* fusion to the aboral ectoderm, the embryonic territory in which *CyIIIa* is normally utilized. P3A2 also binds specifically to similar target sites located in the regulatory region of the *SM50* gene, which is expressed only in skeletogenic mesenchyme lineages. The P3A2 factor was purified by affinity chromatography from nuclear extracts of 24 h sea urchin embryos, and partial peptide sequences were used to isolate a cDNA clone encoding the complete protein. There are no significant similarities between P3A2 and any other protein in existing sequence data bases. P3A2 thus includes a novel type of DNA-binding domain. To examine the differential utilization of P3A2 in *CyIIIa* and *SM50* genes, we measured the specific affinity of this protein for the various target sites in the regulatory DNAs of each gene, and identified the core target site sequences. The stability of P3A2 complexes formed with *SM50* target sites is 50–100 times greater than that of the complexes formed with *CyIIIa* target sites, though the factor binds to very similar core sequence elements. P3A2 is one of at least twelve different proteins whose interaction with *CyIIIa* regulatory DNA is required for correct developmental expression. The results reported demonstrate that it might be possible to purify most of these regulatory proteins, or any other specific DNA-binding proteins of the sea urchin embryo, by using the simple procedures described for P3A2.

Key words. regulatory protein, embryonic gene regulation, *cis*-regulatory target site

Introduction

The early sea urchin embryo develops in a relatively simple manner in which lineages descendant from invariant cleavage-stage founder cells clonally express specific sets of genes. These lineages construct five 'territories,' defined in terms of cell fate and patterns of macromolecular expression (Davidson, 1989). At present the two of these territories for which most molecular data exist are those giving rise to the skeletogenic mesenchyme, and to the aboral ectoderm of the embryo. Founder cells for the skeletogenic mesenchyme are apparently specified autonomously, i.e. by the action of factors present in the cytoplasm of the polar region of the egg, which they inherit, while founder cells for the aboral ectoderm and other territories are apparently specified conditionally, i.e. at least in part by interblastomere interaction, following an initial polarization of the egg cytoplasm in the future oral–aboral axis (Horstadius, 1939; Davidson, 1986, 1989). To approach the mechanisms by which the diverse genetic programs of the polyclonal territories of the embryo are set up, we have undertaken to characterize the regulatory factors that direct expression of marker genes activated specifically in these territories.

This report concerns a factor called P3A2, which appears to participate in regulation of marker genes that are expressed in different territorial domains of the early embryo. P3A2 target sites appear in the regulatory domain of the *CyIIIa* cytoskeletal actin gene, productive expression of which is confined to embryonic and larval aboral ectoderm (Cox *et al.* 1986; Cameron *et al.* 1989; Davidson, 1989). The *CyIIIa cis*-regulatory system includes sites for eleven other DNA-binding factors as well (Calzone *et al.* 1988; Thézé *et al*

1990). Competitive interference *in vivo* with DNA–protein interactions at P3A2 sites and at the target site of another of the *CyIIIa* factors, P7II, causes ectopic expression of a *CyIIIa·CAT* reporter construct (Hough-Evans *et al.* 1990). P3A2 and P7II are thus thought to exert negative control on the *CyIIIa* gene in cells of other territories, *viz* gut, oral ectoderm and mesenchyme. Thiebaud *et al.* (1990) showed that both factors also recognize with high affinity sites present in the upstream region of the *SpecI* gene, another gene expressed exclusively in aboral ectoderm (Lynn *et al.* 1983; Hardin *et al.* 1985; Gan *et al* 1990). However, the P3A2 factor (though not the P7II factor) also binds very tightly to sites located in the control region of a gene called *SM50* (Thiebaud *et al.* 1990). This gene codes for a matrix protein that is expressed exclusively in the lineages of the skeletogenic territory (Benson *et al.* 1987; Sucov *et al.* 1987, 1988). These lineages arise from entirely different founder cells than those specified as aboral ectoderm precursors (Davidson, 1989). The P3A2 factor is encoded by a rare maternal message (Cutting *et al* 1990), and the active factor can be extracted from mid-cleavage nuclei, i.e. it is present during the initial processes of territorial founder cell specification. P3A2 may serve as an element in several different early 'specification switches', in which the sense of the specification is defined by the combination of factors engaged.

A natural advantage of sea urchin embryos for the molecular analysis of territorial gene regulation is the enormous amount of biological material available. We describe herein relatively simple procedures by which several billions of synchronous embryos were harvested, the nuclei separated, and active nuclear extract prepared. Though present at about a thousand molecules per nucleus at the stage from which the extract was obtained (Calzone *et al.* 1988), the P3A2 factor could be purified by site-specific affinity chromatography, and cloned. This interesting factor turns out to be unrelated in sequence to any known DNA-binding proteins. A feature of likely biological interest revealed by measurements carried out with affinity-purified P3A2 is that this protein discriminates sharply amongst the different marker genes, in respect to the stability of its interactions with the various target sites. It follows that different concentrations of P3A2 *in vivo*, in space or time, would be predicted to result in very different patterns of P3A2 regulatory interactions.

Materials and methods

Large-scale embryo cultures

Standard methods for culture of *S. purpuratus* sea urchin embryos are well known Here we address the special problems associated with culture and harvest of unusually large numbers of embryos Eggs from 200–400 gravid *S purpuratus* females were collected over beakers of ice-cold, filtered sea water after intercoelomic injection with 0 5 M KCl. The settled eggs were combined and diluted to a 10–15 % suspension in fresh, ice-cold filtered sea water, and allowed to settle at 4°C in 1 l beakers. One billion settled *S. purpuratus* eggs occupied a volume of about 1 l. The washing procedure was repeated 4–5 times Fertilization efficiency and viability were reduced if the eggs were not washed sufficiently or remained packed too long in larger pellets. The washed eggs were stirred gently in a 10–15 % suspension, while counting an aliquot and making preparations for fertilization The eggs were transferred at a concentration of 20 000–30 000 per ml at 16 °C to 4, plastic 20-gallon containers (Rubbermaid Garbage Cans) in a total volume of about 80 l. The eggs were then fertilized by addition of dry sperm freshly diluted into sea water. The cultures were vigorously aerated and stirred at 60 revs min^{-1} with a large paddle. Embryos cultured beyond the late gastrula stage should be diluted to about 5000–10 000 per ml The embryos were harvested after hatching at the mesenchyme blastula stage (21–24 h), in portions of about $1\ 25\times10^9$. About 60 l of embryo suspension were concentrated by filtration with a 51 micron Nitex filter to about 4 l, transferred to 6 flat-bottomed 1 l bottles, and pelleted by centrifugation at about 1200 *g*. The embryo pellets in each tube were resuspended in about 900 ml of ice-cold 1 M glucose, and collected by centrifugation at 1800 *g*. The washed embryo pellets were then resuspended in about 10× the pellet volume with Buffer A [10 mM Tris (pH 7.4), 1 mM EDTA, 1 mM EGTA, 1 mM Spermidine–Tris–HCl, 1 mM dithiothreitol (DTT), 0 36 M sucrose], and frozen in 1 l aliquots in heavy duty freezer bags (Ziplock) in liquid nitrogen Frozen embryos were stored at −70 °C

Preparation of nuclear protein extracts

Nuclear protein extracts were prepared as described previously (Calzone *et al.* 1988) with several modifications. The frozen embryos were crushed and uniformly thawed in batches of about 2.5×10^9. After mixing with a high speed overhead stirrer, the nuclei were collected by centrifugation in 1 l flat-bottomed bottles (Nalgene), at about 2500 *g* for 40 min at 4°C. The pelleted nuclei were washed by resuspension in about 2–4 l of ice-cold Buffer A and recentrifuged If nuclei trapped in fertilization envelopes were observed in the supernatant, they were released by mixing at moderate speed two times for 30 s in a Waring blender. The Buffer A washes were repeated two times, and were followed by 2–3 washes with Buffer A containing 0 1 % NP-40. By the end of these washes the nuclei were contained in a single 1 l bottle According to Bradford (1976) measurements, the nuclear pellets include about 2 5 % of total starting embryo protein. The nuclei were resuspended in 3.0–3.7× the pellet volume with Buffer D [10 mM Hepes (pH 7.9), 1 mM EDTA, 1 mM EGTA, 1 mM Spermidine–Tris–HCl, 1 mM DTT, and 10 % glycerol] The nuclei were transferred to 35 ml Oak Ridge tubes, and proteins were released from chromatin by addition of one-tenth vol of 4 M ammonium sulfate (pH 7.9), with vigorous mixing, followed by incubation on ice for 1 h, with occasional gentle mixing, and centrifugation at 35 000 revs min^{-1} in Beckman 50 Ti.2, 60 Ti or 70 Ti rotors for 3–6 h at 4°C (Parker and Topol, 1984). Additional supernatant was recovered by recentrifugation of the chromatin pellets for 6 h. Roughly 60–70 % of the total protein in the nuclear pellet typically remained associated with chromatin Approximately 50 % of the solubilized proteins were precipitated from the supernatant by addition of 0.25 gm of ammonium sulfate per ml and incubation overnight on ice These precipitated proteins were collected by centrifugation at 10 000 *g*, resuspended in about 20 ml of Buffer C [20 mM Hepes buffer (pH 7.9); this pH value refers to a 10× stock solution of Buffer C salts, 0.1 mM EDTA, 40 mM KCl, 1 mM DTT, 0.1 % NP-40, 20 % glycerol] The proteins were dialyzed overnight at 4°C against Buffer C. About 40 to 70 % of the protein in the

fraction precipitates at this step, and was removed by centrifugation at 35 000 revs min^{-1} in a 60 Ti rotor for 15 min. The samples were stored at −70°C The final yield of protein was about 10 % of the starting protein in the nuclei.

Site-specific DNA affinity chromatography

Site-specific DNA affinity columns were prepared essentially as described by Kadonaga *et al.* (1987). The oligonucleotides used to prepare the two synthetic binding sites used for purification of P3A2 are as follows For the *CyIIIa* gene, complementary oligonucleotides 11/12 were. (11), GAAGCGCAAACAAACTTTATTAAGC; (12), CTTCGCTTAATAAAGTTTGTTTGCG; for the *SM50* gene, oligonucleotides 25/26 were (25), GCTTCTGCGCACACCCCACGCGCATGGGGCGT; (26), AAGCACGCCCCATGCGCGTGGGGTGTGCGCAG. Columns were stored at 4°C. To purify P3A2, the nuclear protein extracts were applied to site-specific DNA affinity columns without prior purification For each affinity-purification, a quantity of nuclear extract estimated to contain enough factor to saturate approximately 30 % of the total specific sites coupled to the column resin was applied to each column. Application of less extract reduced the ratio of P3A2 to nonspecific proteins in the bound fractions. A 0.2 ml *SM50* column was used for the pilot experiment described in text. A 1.0 ml *CyIIIa* and a 1 5 ml *SM50* columns were used for the large scale purification. The volume of nuclear extracts applied to affinity columns was 3.3 ml for the pilot experiments, and 52 ml for the large scale preparation. When loading large volumes, the extracts were passed through a 0.5 ml Sepharose CL-4B pre-column that did not contain DNA to prevent accumulation of any insoluble debris in the site-specific DNA affinity column. Affinity chromatography was carried out at 4°C Before loading, the concentration of KCl in the nuclear extracts was increased to 0.1 M. All loading, washing and elution buffers contained 20 mM Hepes (pH 7.9), 0.1 mM EDTA, 1 mM DTT, 0.1 % NP-40, 20 % glycerol, and the concentrations of KCl indicated here and in Results. The loading rate was approximately 0 1–0 2 column volumes per min. After loading the column was washed with 30–50 ml of column buffer containing 0.1 M KCl at a rate of 1 ml per min. Bound proteins were eluted with a KCl step gradient (3 column vol per step; an increase of 0.1 M KCl per step) at a rate of about 0.5 column vol per min. For the second cycle of affinity chromatography, the appropriate fractions were pooled (see Results), dialyzed against 0.1 M KCl in column buffer, loaded and eluted as described above

DNA-binding assays

DNAase I footprinting, methylation interference and gel-retardation assays were performed according to standard procedures, which have been described previously (Calzone *et al* 1988, Thézé *et al* 1990) The amount of protein, probe, nonspecific DNA, and DNAase I had to be optimized for each experiment. The *CyIIIa* gene probe in the purification and characterization of purified P3A2 was an 86-bp *Hin*dIII–*Hin*fI fragment of the *CyIIIa* gene regulatory region (−244 to −157), or a *Hin*dIII–*Dde*I fragment (−244 to −114); (see Thézé *et al.* 1990, for the sequence) The *SM50* gene probe for similar experiments was an *Hpa*II–*Bst*NI fragment of the regulatory region (−49 to −200) subcloned into Bluescript KS (Strategene, kindly provided by K. Whittaker) and released from vector by digestion with *Eco*RI and *Hin*dIII A binding site for the CTF transcriptional activator is included in the *Hpa*II–*Bst*NI fragment. The following oligonucleotide gene probes were used to compare the equilibrium binding constants of P3A2 interactions with *SM50* and *Spec*I P3A sites For the *SM50* gene, oligonucleotides 50/51: (*50*), GATCTTTTCGGCTTCTGCGCACACCCCACGCGCATGGGGC, (*51*), GATCGCCCCATGCGCGTGGGGTGTGCGCAGAAGCCGAAAA For the *Spec*I gene, oligonucleotide 52/53. (*52*), GATCATCTGCGCATGCACAGATCAATCCGCGCATGCTCAG; (*53*), GATCCTGAGCATGCGCGGATTGATCTGTGCATGCGCGAT. The procedure for detection of P3A2 DNA-binding activity in nitrocellulose filter blots of one-dimensional SDS protein gels (Laemmli, 1970) was similar to the protocol described by Vinson *et al* (1988) for screening cDNA expression libraries.

Gel purification and renaturation of P3A2

A gel slice containing approximately 1 μg of P3A2 in a SDS protein gel was located using parallel marker lanes, excised, mashed in 0.5 ml of elution buffer [50 mM Tris (pH 7.6), 0 1 mM EDTA, 0 1 % SDS, 5 mM DTT, 150 mM NaCl, 0 1 mM PMSF] and incubated at room temperature for 6 h. Gel pieces were removed by centrifugation and washed twice with 0 2 ml of water. The supernatants were combined and protein was precipitated by addition of 5 vol of acetone, followed by incubation overnight at −20°C. The protein precipitate was collected by centrifugation, washed with fresh acetone–water (5:1), briefly dried and resuspended in 0 1 ml of denaturation buffer [20 mM Hepes (pH 7 9), 0 1 mM EDTA, 1 mM DTT, 6 M guanidinium–HCl, 1 mM $MgCl_2$]. The sample was passed over a 0 5 ml BioRad P6 column equilibrated with 10 mM Hepes (pH 7.9), 0.1 % NP-40, 1 mM DTT, 100 mM KCl, 10 % glycerol. The fractions containing DNA-binding activity were located by the gel retardation method. Compared to the starting material, the efficiency of recovery of specific DNA-binding activity after renaturation was less than 1 %

Protein sequencing

Protein sequencing of affinity-purified P3A2 was accomplished as described by Aebersold *et al* (1989) Approximately 15–30 μg of P3A2, purified by two cycles of chomatography on an *SM50*-P3A site-specific DNA affinity column, were separated from remaining protein contaminants by electrophoresis on a 7 % SDS gel (3–5 μg per lane); and electrophoretically transferred to nitrocellulose (S&S, 0.45 μm). Tryptic fragments of P3A2 separated by HPLC, and sequenced on polybrene-coated glass fiber supports using an Applied Biosystems 477A/120A pulsed liquid sequenator, essentially as recommended by the manufacturer Phenylthiohydantoin-amino acid analysis was performed on-line and sequence assignments were made by visual inspection of the chromatograms The protein sequences obtained are listed below. Numbers within parentheses specify the HPLC fractions from which the peptides came. One-letter codes are used for the amino acids Lower case letters signify uncertain assignments. At position 10 in fraction 29.2, two signals were observed, neither of which could unambiguously be assigned to the major sequence (*14*), ATIDEYATR, (*17 1*), SFETNPSIR; (*17 2*), KVSLA; (*19 1*), VGQQAVV; (*19.2*), LKATIYELVLKkgk, (*19.3*), SSVIND; (*20*), TIVINCYK; (*25*), QTVVAGDGQPIQIANVNIAQQSGxxgTMAAIK, (*29 1*), VFGAAPLENIMR, (*29 2*), SQVLPvFIe(N,R); (*30.1*), GIVLQDLDNASAQrk; (*31.1*), APQPSNENSDSYELPPLVIDGIHDxh; (*31.2*), NAVMQSQPIPLQVATLVVNAASPTQVNK.

Isolation and sequencing of P3A2 cDNA clones

The partial protein sequences of P3A2 were used to construct the oligonucleotide probes shown below (I indicates inosinic acid). (*P25*), CTGCTGIGCAATGTTIACATTIGCAATCTGGATIGGCTGTCCATC; (*P31 1*), GTCATGGATACCAT-

CAATGACIAGTGGTGGIAGCTCATAIGAGTCIGAGT-TCTCATT; (*P31.2*), GTGGCCACCTGIAGIGGGATIG-GCTGIGACTGCATGACAGCATT, (*P30 1*), CTGGGCI-GAIGCATTATCIAGATCCTGIAGGACAATTCC, (*P14*), TGIGC(G,A)TA(T,C)TC(G,A)TCIAT. A set of duplicate filters of a λ ZAP (Stratagene) cDNA library of 14 h embryo poly(A) RNA was reacted with a hybridization mixture containing 0.5 picomole of each oligonucleotide, labeled with ^{32}P at the 5′ terminus, in 6×SET [1×SET is 0 15 M NaCl, 10 mM Tris (pH 8.0), 1 mM EDTA], 50 mM sodium-phosphate (pH 7.4), 5×Denhardt's solution (Denhardt, 1966), 0.1 mg ml^{-1} salmon sperm DNA at 37°C for 18 h The filters were washed twice at room temperature for 10 min each in 2×SSC (1×SSC is 0 15 M NaCl, 15 mM sodium citrate), 0 2 % SDS twice at room temperature in 1×SSC, 0.2 % SDS, twice in 3 M tetramethylammonium chloride (TMA), 50 mM Tris (pH 8.0), 2 mM EDTA, 0.1 % SDS at 37°C for 15 min, and once in the same TMA buffer at 48°C for 15 min. Seven positive clones were detected in a total of 6×10^5 phage. The two longest clones, p21 and p25, were selected for further analysis cDNA inserts were sequenced using standard procedures for the Bluescript system (Strategene).

Preparation of recombinant P3A2

A T7 RNA polymerase based expression system in *E coli* developed by Studier and Moffat (1986) was used to prepare recombinant P3A2. The plasmid construct used to express P3A2 was an *Eco*RV–*Dra*I cDNA fragment, joined using *Bam*HI linker into the *Bam*HI site of pET3c. *Eco*RV cuts in codon 6 of the P3A2 message sequence. The *Dra*I site is located in the 3′ noncoding trailer of the message. The host for P3A2 expression was *E. coli* strain (lys S) To prepare P3A2, a 0.5 l culture of cells containing the expression plasmid was grown to an OD_{600} of 0.7 in NZ medium containing 50 μg ml^{-1} ampicillin and 20 μg ml^{-1} chloramphenicol NZ [10 g l^{-1} NZ amine (Sheffield), and 5 g l^{-1} NaCl]. The T7 RNA polymerase promoter was induced by addition of 0 5 mM IPTG and incubation of the culture was continued for 4 h. The cells were collected by centrifugation, resuspended in 50 ml of fresh NZ medium, pelleted again by centrifugation, and resuspended in 10 ml of 50 mM Tris (pH 7.6), 50 mM NaCl, 1 mM EDTA, frozen in liquid nitrogen, and stored at −70°C. After thawing, addition of 2 mg ml^{-1} lysozyme, incubation on ice for 2 h, and addition of PMSF to 1 mM, lysis was completed by sonication (4 to 5 high intensity bursts) NP-40, sucrose and DTT were then added to concentrations of 0.5 %, 5 % and 1 mM, respectively Insoluble material was removed by centrifugation at 10 000 *g* Greater than 90 % of the recombinant P3A2 was recovered in the soluble supernatant. After addition of one-tenth volume of 4 M ammonium sulfate and glycerol to a final concentration of 20 %, ribosomes and other particles were removed from the supernatant by centrifugation at 100 000 *g*. Recombinant P3A2 constituted at least 1 % of the total protein in the postribosomal supernatant. The proteins were diluted 100- to 1000-fold for DNA-binding protein assays; dialysis to remove ammonium sulfate was not necessary.

Antibody blotting

All procedures were carried out at room temperature. Nitrocellulose blots were incubated for 3 h in 'Blotto' to block nonspecific binding sites (Blotto is 5 % nonfat milk, 0.01 % AntifoamA, 0 0001 % merthiolate in PBS, pH 7 5) Whole rabbit serum containing P3A2 antibodies was diluted 1/400 in 15 ml Blotto and incubated with the blot for 1 h. The primary antiserum was removed and the blot was washed with four changes of Blotto (15 min each). After a 10 min wash in PBS, the blot was incubated for 10 min in 0.1 % glutaraldehyde in PBS followed by two additional washes in PBS (Ikegaki and Kennet, 1989). Secondary antibody, goat anti-rabbit conjugated to horseradish peroxidase (EY Labs), was diluted 1/1000 in 15 ml of Blotto and subsequently incubated for 1 h. The antibody was removed with four 15 min washes of Blotto and the blot was stained with a DAB-metal substrate for 15 min (0.1 % 3,3′ diaminobenzidine, 0.03 % cobalt chloride, 0 05 % H_2O_2 in PBS, pH 7.5). The stained blot was photographed with Kodak Plus-X film

Results

Purification of P3A site binding factors

A sea urchin blastula contains about 5×10^5 molecules of a factor that produces a footprint over the P3A site in the *CyIIIa* and *SM50* genes (Calzone *et al.* 1988; and data shown below in Table 1). The factor was purified by directly applying crude nuclear protein extract from about 2.5×10^9 embryos to site-specific DNA-affinity columns on which were mounted the *CyIIIa* or the *SM50* P3A target binding sites. The specific sites were those located at positions −169 to −193 in the regulatory domain of the *CyIIIa* gene (the sequence is presented by Thézé *et al.* 1990); and the double P3A site at position −98 to −129 in the *SM50* gene (Sucov *et al.* 1988; Thiebaud *et al.* 1990). Details including column oligonucleotide sequences and operating protocols are given in Materials and methods. The elution of specific DNA-binding proteins from each column was monitored by gel-retardation assays and by SDS protein gel electrophoresis. Results from the first rounds of affinity-purification are shown in Fig. 1. P3A activity can be seen to elute from both columns with a protein of approximately 62×10^3 M_r that was easily detected in the silver-stain pattern of the SDS gels. Note that a higher concentration of salt (≥0.7 M KCl) was required for elution of the activity from the *SM50* column than from the *CyIIIa* column (0.3–0.4 M KCl). A consequence was that when purified on the *SM50* column the P3A activity was relatively free of nonspecific proteins (see Table 1). As shown below, a large preference in the specific binding to the *SM50* P3A site over the *CyIIIa* site underlies the differential elution of P3A activity from these two columns. DNAase I footprint and methylation interference patterns obtained after reaction of the proteins purified on the *SM50* column are shown in Fig. 2. These patterns are identical to those obtained with crude extracts, using the same double P3A target site of the *SM50* gene (data not shown).

The protein preparation obtained after two cycles of affinity chromatography with the *SM50* column (Fig. 1B) contained two major species, having apparent relative molecular masses of 62 and 110×10^3, respectively, though as can be seen in Fig. 1A the 110×10^3 species is also present in column fractions that display no P3A activity. In Fig. 3 we show experiments demonstrating that the specific DNA-binding activity recovered from the *SM50* column is entirely due to the 62×10^3 protein, hereafter called P3A2. Thus the 62×10^3 protein reacted with a probe containing

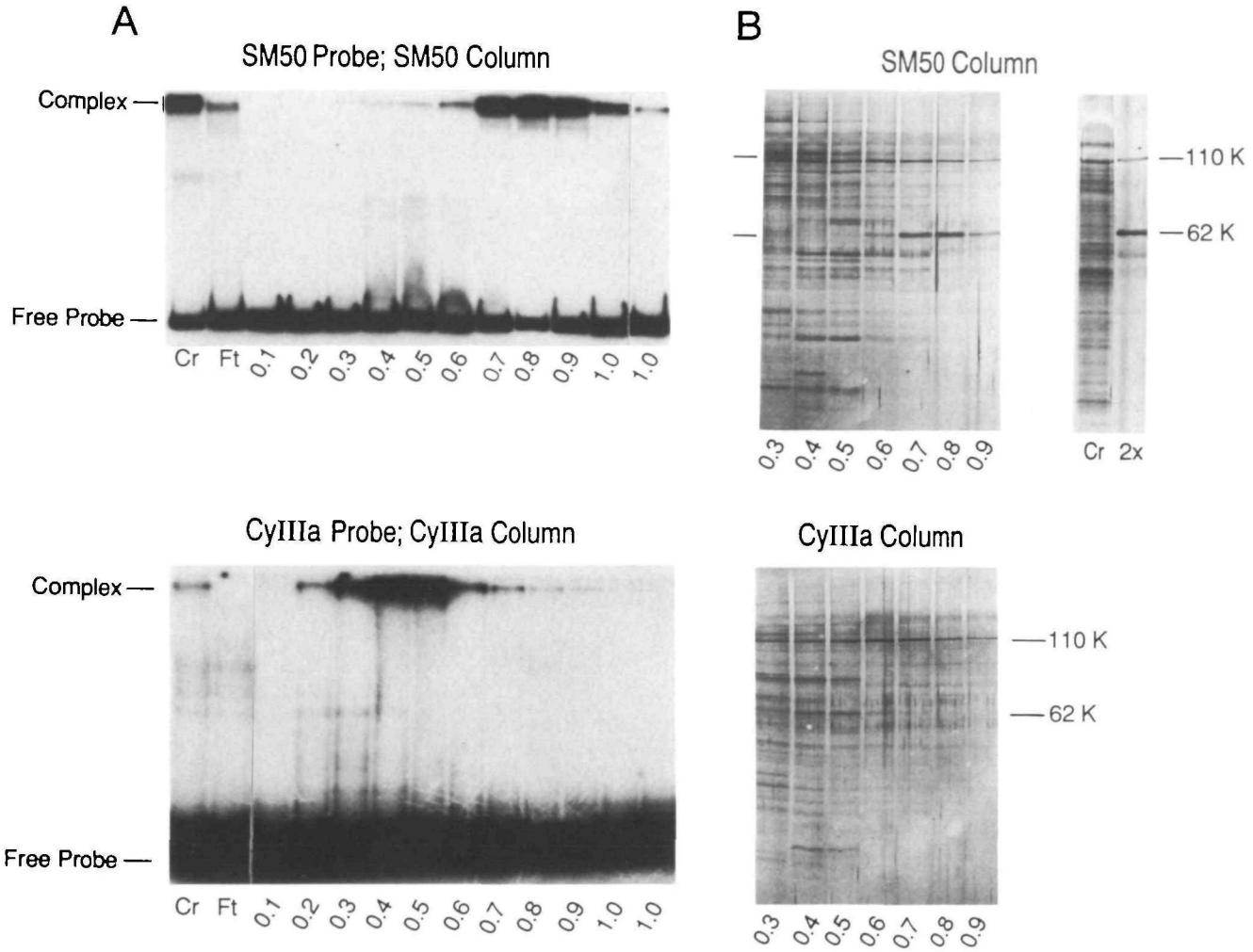


Fig. 1. Affinity purification of P3A2. Blastula nuclear extracts representing about 2.5×10^9 embryos (10^{12} nuclei) were passed sequentially over a *CyIIIa* and a *SM50* site-specific column (see Materials and methods for details). After washing extensively the bound proteins were eluted with a step gradient of KCl. Approximately 7×10^{-9} moles of specific factor were applied to the affinity columns The *CyIIIa* column bound about 50% of the P3A2 in the extract The majority (90%) of the remaining factor was bound by the *SM50* column (it is likely that the capacity of the *CyIIIa* column for P3A2 was exceeded by a factor of about two). The elution of bound proteins was monitored by gel retardation assay or SDS gel electrophoresis as shown, the concentration of KCl in the bound fractions assayed is indicated below each lane. (A) Gel retardation assays. The probe used to monitor the elution of specific proteins from the *SM50* column was the *Hpa*II–*Bst*NI fragment of the *SM50* regulatory DNA shown in Fig 2. Protein binding to the *CyIIIa* column was assayed with the *Hin*dIII–*Dde*I fragment of the *CyIIIa* gene shown in Fig. 5. The binding reactions (10 μl) contained $8\ 3\times10^{-11}$ M of the *SM50* probe or 6.8×10^{-11} M of the *CyIIIa* probe. Lanes labeled 'Cr' contain crude, or unfractionated nuclear extract, and lanes marked 'Ft' contain samples of the flow-through fractions from the affinity columns. For Cr and Ft reactions 8.8 μg of poly (dI/dC) (*SM50* probe) or 5.6 μg of poly (dAT) (*CyIIIa* probe) were added Appoximately 10^{-5} of the total protein in the Cr and Ft fractions was used for a given assay. The amount of poly (dI/dC) or poly (dAT) in reactions carried out with higher KCl fractions eluted from the affinity columns was reduced by 70%. About 10^{-4} of the total protein recovered in each KCl fraction was used per reaction. (B) SDS protein gels. Approximately 3 μl of each column fraction indicated was analyzed by electrophoresis on a 10% SDS–polyacrylamide gel Proteins were detected by silver stain. The proteins present in the preparation of P3A2 obtained after two cycles of affinity-purification with an *SM50* column are shown in the *SM50* panel, lane 9, marked '2x'

concatenated P3A binding sites (oligonucleotides 25/26) in protein gel blots, but not with a probe containing *CyIIIa* binding sites for a factor we call P1 (Fig. 3A). The same negative results (not shown) were obtained with probes representing binding sites for other factors, P6 and P7I. However, the 110×10^3 protein reacted equally well with all of these unrelated binding sites, and thus is to be considered a nonspecific DNA-binding protein. The 62×10^3 protein was gel purified, renatured, and shown to form a specific complex with a probe containing the *SM50* P3A site by the gel shift method (Fig. 3C), and by DNAase I footprinting (Fig. 3B). As Fig. 3C shows, the 110×10^3 protein did not reveal any specific DNA-binding activity.

Table 1. *Pilot scale purification of P3A2 from blastula nuclear extracts*

(1) Fraction	(2) Factor[a] (moles)	(3) Total protein[b] (g)	(4) Sp Act [c] (moles/g protein)	(5) Recovery[d] (%)	(6) Purification[e] (fold)
Nuclear extract	3.0×10^{-10}	9.2×10^{-2}	3.3×10^{-9}	100	1
First *SM50* site column[f] (0.7–1.0 M KCl)	2.3×10^{-10}	2×10^{-3}	1.1×10^{-7}	76	33
Second *SM50* site column (0.5–0.9 M KCl)	1.6×10^{-10}	3×10^{-5}	5.3×10^{-6}	54	1600

[a] The amount of P3A2 factor in blastula nuclear extracts was determined as described previously (Calzone *et al* 1988, see Fig 7 of this paper) using the *SM50* probe shown in Fig 2, oligonucleotide 25/26 concatenate as a specific competitor, and poly dAT as a nonspecific competitor The oligonucleotide 25/26 competitor contains the tandem P3A sites in the *SM50* gene used to prepare the *SM50* affinity column (see Materials and methods for details and Fig 2) The extract used in the purification represented 1.6×10^{8} embryos (6.3×10^{10} nuclei)

[b] The amount of protein in each fraction was determined by the method of Bradford (1976)

[c] The specific activity of P3A2 at each step was calculated as [moles of factor (column(2)/total protein (3))]

[d] Recovery was calculated as [moles of factor (2)/3.3×10^{-10}]

[e] Purification was calculated as [specific activity (4)/3.3×10^{-9}].

The information obtained in the pilot purification summarized in Table 1 allowed us to estimate the relative concentration of P3A2 in blastula nuclear extracts in two different ways. A conservative measure of the purity of the P3A2 preparation, after two cycles of affinity chromatography with an *SM50* column, based on the silver-stained SDS gels, indicated that P3A2 constitutes about 30 % of the total protein (30 μg) in the bound fraction. The final efficiency of recovery of P3A2 was about 54 %, and thus the total amount of P3A2 in the crude extract was approximately 17 μg ($(0.30\times30\ \mu g)/0.54$). Second, according to a gel retardation measurement, the quantity of nuclear extract used in the pilot purification of Table1 contained about 300 picomoles of the factor that interacts with the P3A probes. Assuming the protein binds specific DNA as a monomer and using 49×10^{3} as the relative molecular mass for the factor (established by sequencing the cloned P3A2 mRNA; see below), we calculated that the crude extracts of blastula nuclei contained about 24 μg of P3A2. The good agreement of these two different estimates of P3A2 yields achieved in the pilot purification should not be taken as proof that P3A2 binds DNA in monomeric form. Thus differential silver-staining of the nonspecific proteins in the affinity-purified preparation could easily have led us to underestimate the purity of P3A2 by a factor of two The mass ratio of P3A2 to total blastula nuclear extracts was estimated to be about 2×10^{-4} (20 μg/92 mg; see Table 1). Of the total enrichment required for purification of P3A2 from the crude nuclear pellet, a factor of about 10 was achieved by elimination of 90 % of the total protein in the nuclear pellet by the various steps in the preparation of the nuclear extracts (see Materials and methods), and the remainder was accomplished by affinity chromatography. The results of the P3A2 purification suggest that the isolation of other DNA-binding factors of similar prevalence should be relatively straightforward by the same procedures. Aebersold *et al.* (1988) have shown that for the purpose of protein sequencing a purity of about 5 % is generally sufficient prior to protein purification on SDS protein gels. Thus, a relatively modest 250-fold purification would have been adequate for P3A2. The concentrations of the other factors in blastula stage nuclear extracts that have been shown to bind specifically to the *CyIIIa* regulatory DNA (Calzone *et al* 1988) range from about 0.2 to 4 times the concentration of P3A2 (1×10^{5} to 2×10^{6} molecules/400 cell embryo). Thus the minimum purification required for each of these factors would range from 62- to 1250-fold. These are levels of purification easily achieved by direct site-specific DNA affinity chromatography of the crude nuclear extracts.

Cloning P3A2 mRNA

Approximately 0.5 nanomole of the 62×10^{3} P3A2 protein obtained with the *SM50* column was further purified by SDS–polyacrylamide gel electrophoresis (see Fig. 1), transferred to a nitrocellulose filter, and excised and digested *in situ* with trypsin. Peptides released from the filter were separated by HPLC and protein sequences obtained; these are shown in Materials and methods. These sequences were used to construct five oligonucleotide probes, also given in Materials and methods, with which clones encoding P3A2 were recovered from a cDNA library representing 14 h embryo poly(A) RNA. Seven cDNA clones were isolated from a total of about 6×10^{5} recombinant phage. The longest insert (clone p21) was 3664 nt, close to the full length of the P3A2 message according to RNA gel blot measurements (Cutting *et al.* 1990). Alignment of the tryptic peptide sequences of the purified P3A2 with the predicted protein encoded in the p21 cDNA insert revealed the peptide sequences, within an open reading frame beginning with an ATG at position 79, and ending at a stop codon at position 1446. The complete sequence is shown in Fig. 4, where the location of each tryptic fragment sequence in indicated. The derived protein is 49×10^{3}, i.e. significantly smaller than the estimate of 62×10^{3} provided by protein gel electrophoresis (Fig. 1). To confirm that clone p21 indeed encodes P3A2, a recombinant protein contain-

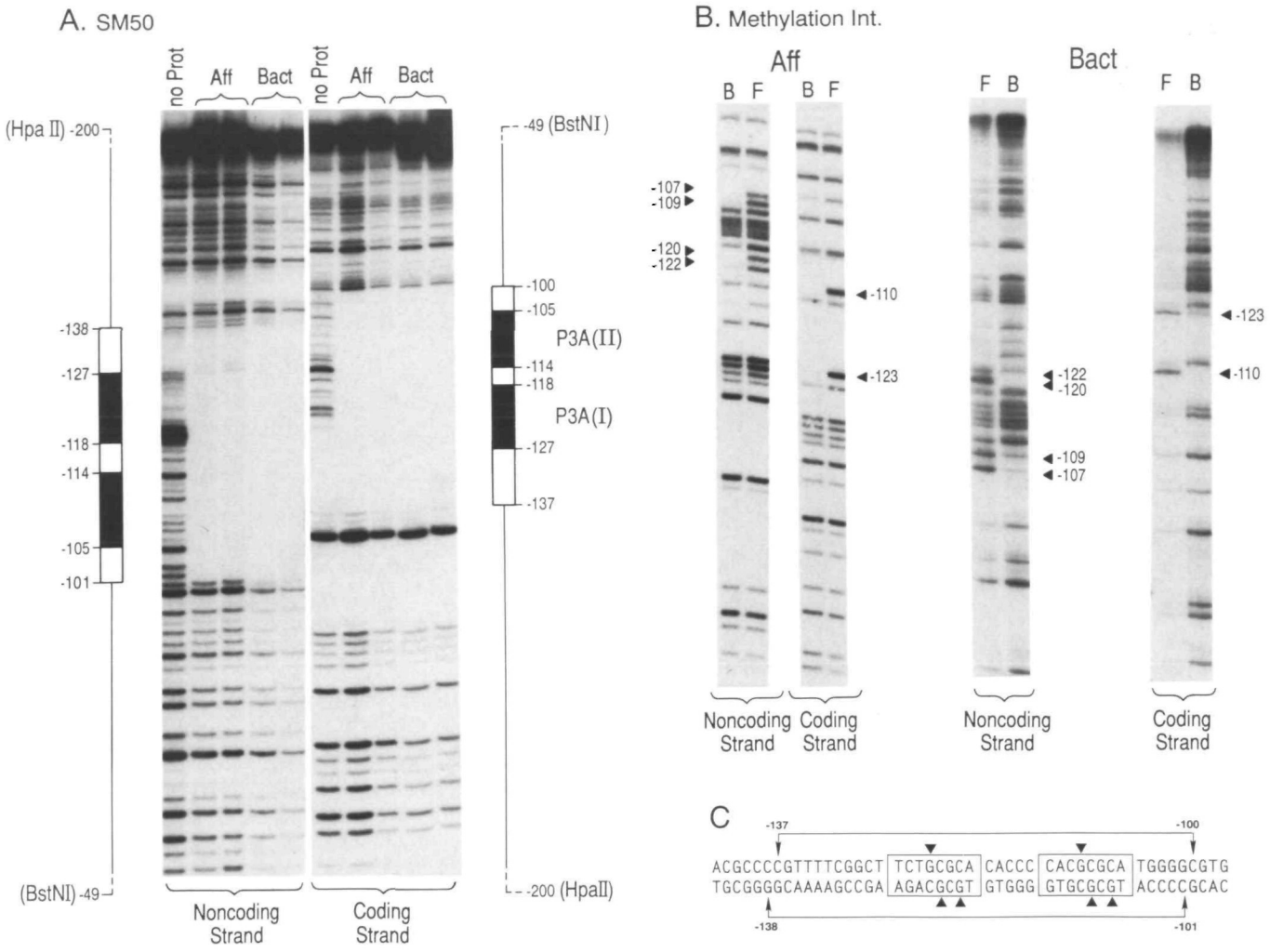


Fig. 2. DNAase I footprint and methylation interference reactions of P3A2 with *SM50* probes. (A) DNAase I footprints. Probes labeled by the kinase reaction at the *Bst*NI (noncoding) or *Hpa*II (coding) terminus were reacted with (left to right) 0 μl (no Prot), 1.0, 2 0 μl (Aff) of P3A2 purified by two cycles of affinity chromatography (first preparation, see Table 1) or 1 0, 2.0 μl of bacterial P3A2 extract and treated with 60 μg ml^{-1} of DNAase I for 1 min on ice The binding reactions (12.5 μl) contained 5 μg of poly (dAT) and about 0.3 ng of specific probe The preparation of the recombinant P3A2 is described in Materials and methods. Marker lanes are not shown The regions of the probes protected from nuclease attack by the P3A2 proteins are represented by the boxes. The filled boxes within the protected sequences map the position of P3A core target site sequences. (B) Methylation interference assays. Noncoding and coding probes were separately treated with DMS and reacted with affinity-purified (Aff) or bacterial (Bact) P3A2. Specific complexes and free probe were separated by gel electrophoresis, eluted and cleaved at residues with piperidine. The cleavage products for each probe derived from specific complex are in lane B, free probe products are in lane F The solid triangles identify the sites of strong methylation interference in the binding of P3A2 with each strand-specific probe. The probe for the reactions with affinity-purified protein was labeled by the end-fill reaction at the *Hpa*II end (noncoding) or *Bst*NI end (coding). Reactions with bacterial protein used the same probes shown in (A). (C) Sequence map of DNAase I footprint and methylation interference patterns. The sequence in the region of *SM50* regulatory DNA which was found to specifically interact with affinity-purified P3A2 is shown. The DNAase I footprints on each strand are represented by the brackets. The solid triangles locate positions of strong methylation interference. The boxes map the tandem P3A target site core sequences.

ing amino acid residues 7 to 459 (the C terminus), fused to a short length of vector and linker N-terminal residues, was expressed in *E.coli*, and assayed for DNA-binding activity. The DNAase I footprint and the methylation interference patterns obtained for the recombinant protein with an *SM50* probe were found to be identical to those observed for affinity-purified P3A2. These results are included in Fig. 2.

Characteristics of the P3A2 protein sequence

A search of current data bases revealed no significant similarity between P3A2 and any previously reported protein sequence. A low resolution N-terminal and C-terminal deletion analysis (see accompanying paper, Höög *et al.* 1991) has shown that the sequences essential for the DNA-binding activity of P3A2 span a broad region, the N terminus of which occurs between amino

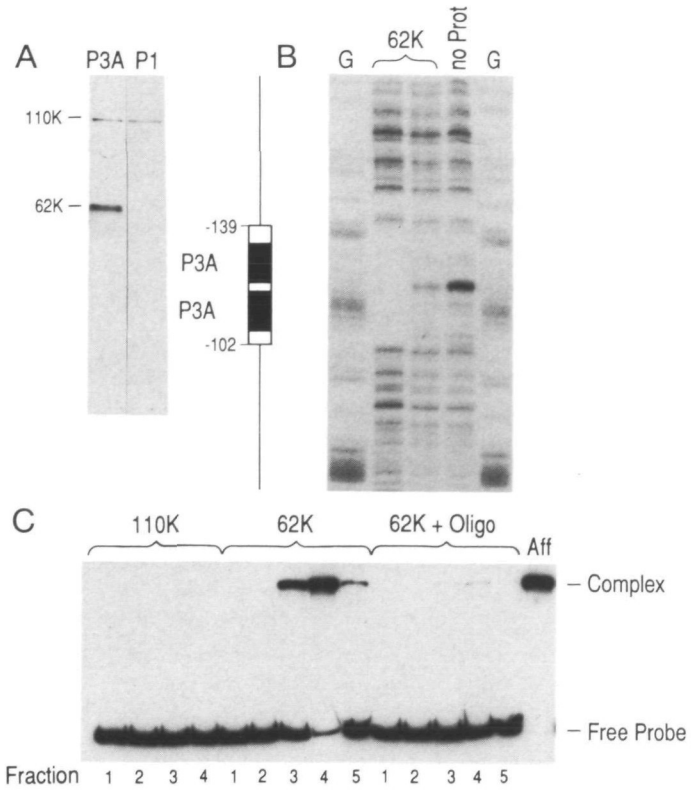


Fig. 3. DNA-binding reactions with gel eluted, affinity-purified proteins. (A) Protein gel blot reactions. Proteins purified by two cycles with an *SM50* column (second preparation, Table 1) were separated by electrophoresis in an 8% SDS–polyacrylamide gel, transferred to nitrocellulose and reacted separately with labeled P3A or nonspecific P1 binding sites (see Materials and methods for details). The P3A probe was a concatenate of oligonucleotides 25/26; The P1 probe was the concatenated binding site 37/38 (Thézé *et al.* 1990). Other oligonucleotide sites tested which generated a reaction pattern exactly similar to P1 (not shown here) were P6 and P7I (oligonucleotides 5/6 and 19/20 of Thézé *et al* 1990) (B) DNAase I footprints The 62×10^3 protein was gel-purified and renatured as described in Materials and methods. The renatured protein (80 μl and 40 μl, left to right) was reacted with a coding-strand probe including the whole *SM50* regulatory region and treated with DNAase I. G indicates a G ladder sequence reaction. (C) Gel retardation assays. Gel-purified 62×10^3 and 110×10^3 proteins were assayed for DNA-binding activity by the gel retardation method using an *SM50* probe containing the tandem P3A target sites (see Fig. 2). The specificity of the complex detected with the 62×10^3 protein was demonstrated by competition with concatenated binding site 25/26 (lanes marked 62 k+oligonucleotide). The lane marked Aff shows the complex present in affinity-purified P3A2 before gel-purification. 'Fraction' refers to eluates from the Biogel P6 column used to remove guanidium HCl (see Materials and methods).

acid positions 25–90, and the C-terminal boundary between residues 222–358, i.e. a DNA-binding domain of at least 130 amino acids is necessary. Nor has the sequence of this DNA-binding domain *per se* yet revealed similarity to any previously characterized DNA-binding motif. P3A2 is thus a novel, sequence-specific recognition factor. Two regions of P3A2 are relatively rich in serine and threonine, and may include sites of phosphorylation detected in the affinity-purified protein (F.Calzone and M. Harrington, unpublished observations). Near the N terminus, the sequence spanning residues 4 to 28 is 36% serine plus threonine (S+T), residues 266–290 are 28% (S+T) and residues 364–389 are 26.9% (S+T). For comparison, the remaining sequences in P3A2 average 11.7% (S+T). P3A2 also has a very glutamine-rich region, similar to those that have been detected in several other DNA-binding proteins, e.g. Oct-1 and Zeste (Pirrotta *et al.* 1987; Sturm *et al.* 1988). Thus the P3A2 sequence between residues 308 and 336 is 37.9% glutamine, compared to 6.7% glutamine in the remainder of the protein.

P3A2 interacts with CyIIIa *regulatory DNA*

Several lines of evidence indicate that P3A2 recognizes regulatory sites in the *CyIIIa* gene, as well as in the *SM50* gene. The *CyIIIa* site reproduced in oligonucleotides 11/12 (sequence given in Materials and methods) and mounted on the *CyIIIa* affinity column is also protected from DNAase I when reacted with P3A2 purified with the *SM50* column. This footprint is shown in Fig. 5A, and Fig. 5B demonstrates that the same essential G residues that are contacted by P3A2 in the *SM50* P3A target site are contacted in the *CyIIIa* target site used for these experiments, although additional G contacts were detected outside of the core sequence. These data are summarized in Fig. 5C, which should be compared with Fig. 2C. In addition, P3A2 factor purified using the *CyIIIa* affinity column produced the same DNAase I footprints and methylation interference patterns noted for P3A2 purified on the *SM50* column (not shown). Finally, as shown in Fig. 6, an antibody raised against *recombinant* P3A2 detected protein of apparent mass 62×10^3 in the specific protein fractions eluted from both the *CyIIIa* and *SM50* columns. Taken together with results of the gene transfer studies summarized in Introduction, these observations suggest that P3A2 participates in regulation of mutually exclusive patterns of gene transcription, in developmentally unrelated cell types, as exemplified by the territory-specific *CyIIIa* and *SM50* genes.

The organization of P3A2 target sites in the CyIIIa *and* SM50 *genes*

DNAase I footprint and gel retardation assays performed with affinity-purified P3A2 revealed multiple binding sites for the protein in the extensive regulatory domains of both the *CyIIIa* and *SM50* genes (data not shown). The coordinates of each site are listed in Table 2. The *CyIIIa* regulatory region contains a total of three P3A2 binding sites. The binding site at position −182 to −195 includes the consensus P3A target sequence identified by Thiebaud *et al.* (1990), i.e. $^{C}/_{T}X^{C}/_{T}GCGC^{A}/_{T}$, while those at positions −807 to −820, and −101 to −114 differ respectively in two and

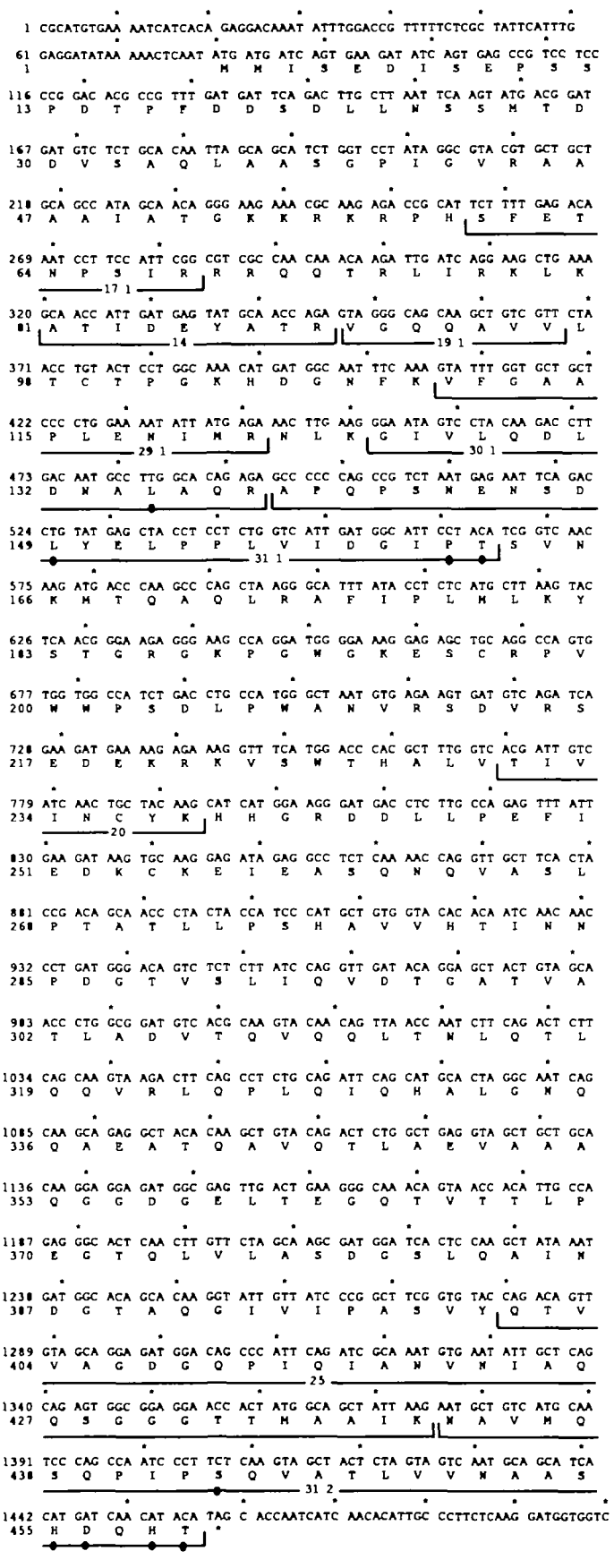
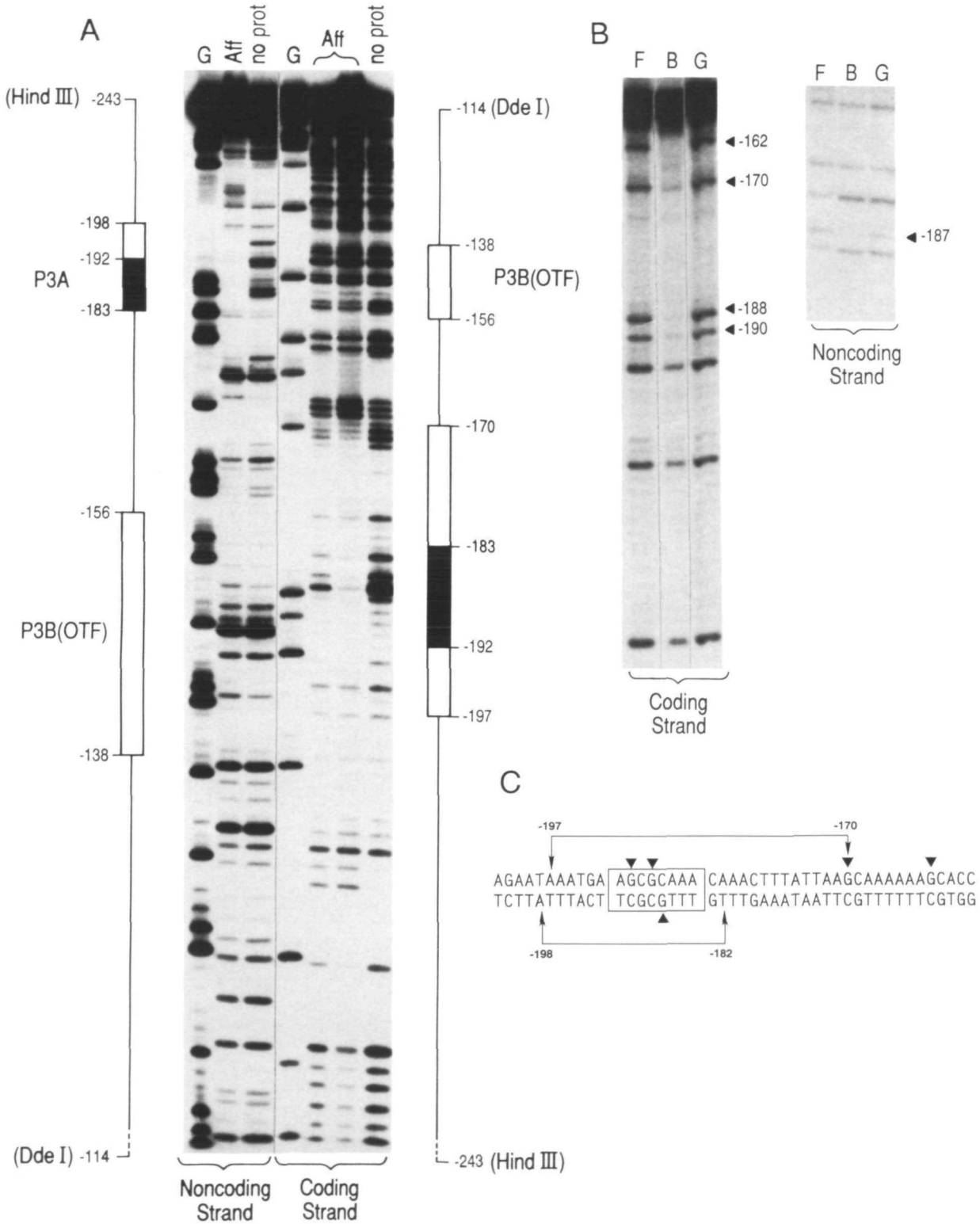


Fig. 4. Sequence of P3A2 mRNA, and of encoded protein. The sequence may be missing up to about 50 nt from the 5′ leader (see text). Brackets indicate locations of predicted peptides obtained by direct sequencing of tryptic peptides from affinity-purified P3A2, as described. All peptides shown in Materials and methods are identified and are indicated, except for several that are redundant, i.e. included within the sequences bracketed. Agreement between the sequenced peptides and the translated sequence of the cDNA was perfect, except where marked with dots. The misidentifications occasionally occurred as confusions between S and L, and at the C-terminal ends of the peptides where the sequence is more difficult to interpret. The sequence has been entered in the Genbank/EMBL database, accession no. X57014.



three nucleotides out of the seven constrained positions (marked by dots in Table 2). Correspondingly, the interaction of P3A2 with the site at −101 to −114 is relatively weak (Table 2), and can only be observed after occupation of the strong site at −182 to −195, and only if both sites are present on the probe fragment. Five P3A sites occur in the regulatory region of the *SM50* gene. The two closely spaced sites at positions −102 to −128 (Table 2) are those that were utilized for the purification of P3A2 described above. As shown in Fig. 2, the pattern of essential G contacts in these sites indicates that P3A2 binds similarly to both of them. Another strong P3A site occurs in the *SM50* regulatory region close to the transcription start, at position −23 to

Fig. 5. P3A2 interactions with *CyIIIa* regulatory DNA. (A) DNAase I footprints The *CyIIIa* regulatory DNA fragment indicated in the Figure was labeled by the polynucleotide kinase method, and reacted with P3A2 obtained from an *SM50* affinity column ('Aff' on Figure); or with no protein, as indicated 'G' indicates marker G ladder sequencing reactions. For the noncoding strand reaction shown, 10 μl of the P3A2 preparation were used, and for the coding strand 8 μl and 2 μl were used. The reaction mixtures were treated with 80 μg ml^{-1} DNAase I for 40 s on ice. The reactions (12 5 μl) contained about 0.015 picomole of each probe. This *CyIIIa* probe also contains an octamer binding site indicated on the Figure ('P3B (OTF)'). (B) Methylation interference assays The *CyIIIa* probes were treated with DMS and separately reacted with P3A2 purified on an *SM50* column. Free probe ('F') was separated from probe bound to P3A2 ('B') by polyacrylamide gel electrophoresis. Both probe fractions were eluted and cleaved with piperidine (C) Sequence map of P3A2 target site at positions −198 to −170 in the *CyIIIa* gene. The sequence in the region of the probes covered by DNAase I footprints in A is shown. DNAase I footprints on each strand are indicated by boxes. The shaded region of each box covers the P3A core target site sequence. Solid triangles mark positions of strong methylation interference, from the experiments shown in B.

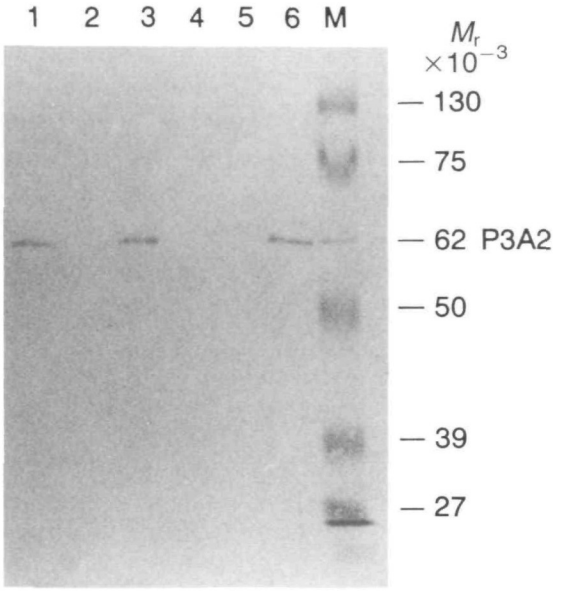


Fig. 6. Identification of P3A2 by reaction with rabbit antibody against recombinant P3A2 protein. Proteins were displayed by gel electrophoresis, transferred to filter paper, and reacted with the antibody, as described in Materials and methods. The control is shown in lane M, where the recombinant P3A2 protein used as antigen was mixed with the size markers. The reactive band is indicated. Lane 1, crude nuclear extract, prior to fractionation, lane 2, flow through fraction from *SM50* affinity column; lane 3, high salt eluate from *SM50* column, and lane 4, preceding low salt eluate from the same column (cf. Fig. 1); lane 5, high salt eluate from *CyIIIa* column and lane 6, preceding low salt eluate from the same column. Note that P3A2 is released from the *SM50* column only in the high salt fraction, but from the *CyIIIa* column in the low salt fraction, a result consistent with the observations reproduced in Fig 1, and also with the equilibrium dissociation measurements described in text. The sensitivity of this assay is limited (compared, e g to gel shift reactions), and residual small amounts of P3A2 could have been present in lanes where no signal is observed

−36, and two other sites, also agreeing perfectly with the consensus, are identified further upstream (see Table 2). If P3A2 also acts in the *SM50* gene as a negative regulator, as the evidence suggests that it does for the *CyIIIa* gene (Hough-Evans *et al.* 1990; Hoog *et al.* 1991; Davidson, 1989), it may function by antagonizing the effects of a variety of *different* positive regulators. Thus P3A2 target sites are juxtaposed to the recognition sites of several transcriptional activators in the regulatory domains of both *CyIIIa* and *SM50* genes (see Thézé *et al.* 1990 and Thiebaud *et al.* 1990). Both regulatory domains contain consensus sites for the positive transcription factor CTF/NF-1 (in these references designated P4), that are in each case located close to a P3A2 target site. In addition, P3A2 sites in the *CyIIIa* regulatory region flank a recognition sequence for an octamer protein (designated P3B), and sites for other positively acting transcription factors in the *CyIIIa* domain including the enhancer factors P2I, P2II, and the temporal activator P5, are also adjacent to the P3A sites listed in Table 2.

Specific binding affinities of P3A2

One path that we have taken to investigate differential utilization of P3A2 by the *CyIIIa* and *SM50* genes is to examine the specific affinity of the protein for the various target sites in each regulatory DNA. In the experiment shown in Fig. 7, fixed amounts of affinity-purified P3A2 and labeled *SM50* probe containing the tandem P3A sites were mixed with increasing quantities of concatenated *SM50* (oligonucleotides 25/26) or *CyIIIa* (oligonucleotides 11/12) target site DNA. The probe bound in specific complex was separated from the free probe by gel electrophoresis, and the ratio of bound to free probe was used to estimate the concentrations of P3A target sites engaged in specific complex, and remaining unbound. The value for the equilibrium dissociation constant, K_D, of the interaction was extracted from these gel shift competition data as described in Fig. 7 and by Hoog *et al.* (1991). An indistinguishable measurement was obtained with recombinant P3A2 so it is unlikely that any minor contaminant present in the preparation could have affected the determination. Fig. 7A shows that the K_D value for the interaction of P3A2 with the *SM50* probe (i.e. the sites shown in Fig. 2C) was 1.7×10^{-9} M. We then measured the *relative* affinity of P3A2 for the various *CyIIIa* and *SM50* target sites, as described in the legend to Fig. 7, and these results, for all the known P3A2 sites in the *SM50* and *CyIIIa* genes, are listed in Table 2. The example shown in Fig. 7B indicates that the affinity of P3A2 for the double sites included in the *SM50* probe is 58× the affinity of this factor for the *CyIIIa* sites included in the standard competitor oligonucleotides (11/12), i.e. the *CyIIIa* site at pos-

Table 2. *Relative affinities of P3A2 target sites*

Gene[a]	Sequence[b]	Relative equilibrium constant[c]
SM50	-128 CTTCTGCGCACACCCACGCGCATGGG -102 GAAGACGCGTGTGGGGTGCGCGTACCC	43 (58)[d]
Spec 1	-84 TATCTGCGCATGCACAGATCAATCCGCGCATGCT -51 ATAGACGCGTACGTGTCTAGTTAGGCGCGTACGA	35[e]
SM50	-23 CCCTCGCGCAACAG -36 GGGAGCGCCTTGTC	8-12[f]
SM50	-356 GTCCCGCGCACACG -343 CAGGGCGCGTGTGC	1-2[f]
CyIIIa	-182 TGTTTGCGCTTCAT -195 ACAAACGCGAAGTA	1.0[g]
CyIIIa	-820 CGGCGGCGCACAAA -807 GCCGCCGCGTGTGT	1-2[h]
SM50	-249 ATTATGCGCTCATC -236 TAATACGCGAGTAG	0.5-1[f]
CyIIIa	-114 CTGAGGCGTACGAT -101 GACTCCGCATGCTA	<0.5[i]

[a] Sites listed are located within regulatory domains that according to gene transfer experiments are necessary and sufficient to promote correct expression of reporter genes to which these domains had been fused. Correct spatial expression has thus been demonstrated for *CyIIIa* gene fusions (Hough-Evans *et al* 1987, 1988, 1990), and for *SM50* gene fusions (Sucov *et al* 1988). Correct temporal expression has been demonstrated for the *Spec*I gene (Gan *et al* 1990), and for the *CyIIIa* gene (Flytzanis *et al* 1987, D. Livant, B. Haigh-Evans and E. Davidson, unpublished results). Sequences of the relevant regions of these regulatory domains have been published for *Spec*I (Hardin *et al* 1985), for *CyIIIa* (Thézé *et al* 1990), and for *SM50* (Sucov *et al* 1988).

[b] Symbolism in the target sequences is as follows: The shaded boxes indicate the degenerate consensus version of the canonical P3A2 core target site, *viz* 5′ $^{C}/_{T}X^{C}/_{T}GCGC^{A}/_{T}$ (Thiebaud *et al* 1990). Regions shaded agree perfectly with this degenerate consensus. The larger open box seen around the proximal *SM50* target site shown in the first row of the Table indicates a site composed of two half sites that are nearly perfect inverse repeats of one another. Note that as displayed some sites have been reversed from their natural orientation with respect to the transcription start site, as indicated by the sequence coordinates, so that the core site is always shown in the orientation given earlier in this Note.

[c] 'Relative equilibrium constant' is calculated as the measured equilibrium constant normalized to the equilibrium constant for the *CyIIIa* site at −182 to −195, i.e. $1.35\times10^{7}\,\text{M}^{-1}$. K_D, *equilibrium dissociation constant*, is reported in text and in Fig. 7, *equilibrium constant* is K_D^{-1}. All measurements shown were obtained using affinity-purified P3A2.

[d] The value shown is from Fig. 7A. The relative equilibrium constant obtained using the *SM50* probe and *CyIIIa*-specific competitor (oligonucleotide 11/12) (Fig. 7B) is given in parentheses.

[e] The probe in this measurement was a 55 bp *Spec*I target site constructed from oligonucleotide 52/53 (see Materials and methods). The specific competitor was the *SM50* oligonucleotide 25/26 concatenate. To control for the large difference in the lengths of the oligonucleotide 52/53 probe and the standard *SM50* probe (−50 to −201), a competition series was carried out with a 55 bp *SM50* probe (oligonucleotide 50/51).

[f] These measurements were obtained by determining the concentration of P3A2 required to occupy about 50% of the target sites in a probe containing the complete *SM50* regulatory region (Sucov *et al* 1988). The reactions were assayed by DNAase I footprinting.

[g] The values shown were taken from an experiment similar to that shown in Fig. 7A. The probe was the 86 bp Z fragment of the *CyIIIa* gene (Thézé *et al* 1990, this fragment contains no other factor binding site). This measurement yielded a value of $K_D=7.4\times10^{-8}$ M, using the oligonucleotide 11/12 concatenate as a specific homologous competitor.

[h] The probe used for this measurement, by the method of Fig. 7B, was fragment N of the *CyIIIa* regulatory domain (see Thézé *et al* 1990), the specific competitor was concatenated *SM50* oligonucleotide 25/26.

[i] Reaction of P3A2 to this target was not observed unless the probe also contained the target site at position −182 to −195. The measurement was carried out by the method described in f using probe F of Thézé *et al* (1990).

itions −182 to −195 shown in Fig. 5C. This result agrees reasonably with that obtained by direct estimation and comparison of the individual K_D values measured by the method illustrated in Fig. 7A. Thus for the same *CyIIIa* site, $K_D=7.4\times10^{-8}$, yielding a ratio of 43 for the comparison with the *SM50* sites assessed in Fig. 7A. Such comparisons represent the *observed* relative affinities, and are thus probably the relevant measure for considerations of differential P3A2 function *in vivo*. Note, however, that when the probes used contained multiple target sites the observed K_Ds may not directly indicate the *intrinsic* stabilities of the respective DNA–protein complexes. We have excluded the possibility that P3A2 binds cooperatively to the two *SM50* sites at position −102 to −128 (see Table 2) since a protein titration (not shown) provides no indication whatsoever of cooperative interaction. Nor is interaction of the two P3A2 proteins that would be bound to these two sites implied by the spacing of the sites, i.e. 14 bp, or 1.33 helical turns. Table 2 also shows a high relative equilibrium constant, about 35× that for the standard *CyIIIa* site, for the interaction of P3A2 with a second double site that is found in the *Spec*I gene. In agreement with this measurement, the results of Thiebaud *et al.* (1990) also indicated a relatively strong reaction between P3A2 (in crude extracts) and the two *Spec*I proximal sites, which are separated by 21 bp. It should be noted that there is as yet no evidence that these sites are functional and they are at least not sufficient to promote accurate expression of a *Spec*I fusion gene (Gan *et al.* 1990). Table 2 also shows that the *SM50* site at position −23 to −36 binds P3A2 about 10× more tightly than does the *CyIIIa* site used for comparison (Table 2). This is not a double site, but it differs from the majority of sites in Table 2 that also agree perfectly with the P3A2 consensus target site sequence, in that it shares three additional nucleotides with one of the strong *SM50* sites, indicated by 'x' in Table 2. Overall the strength of the P3A2 interactions varied about 100-fold. The weakest binding detected was about 50% of the value of the interaction with the standard *CyIIIa* site at position −182 to −195, while all the *CyIIIa* and *SM50* sites *other* than those just discussed bind the factor 1–2× as well as does the standard site. It follows from these measurements that

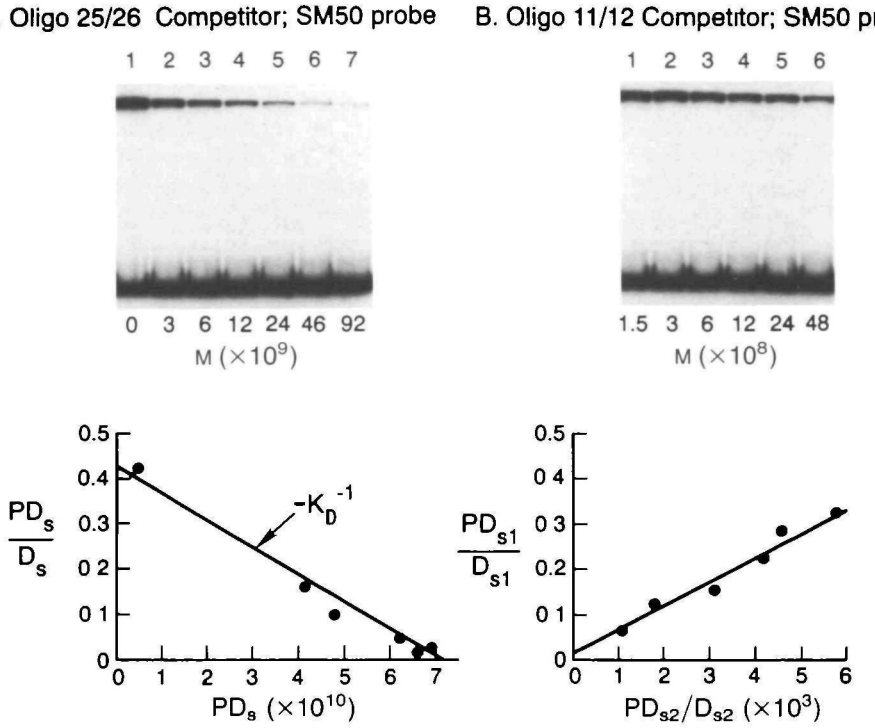


Fig. 7. Gel shift comparisons of P3A2 binding to *CyIIIa* and *SM50* P3A target sites. Representative gel retardation assays used to measure the equilibrium dissociation constants (K_D) for the interaction of affinity-purified P3A2 with *CyIIIa* and *SM50* P3A target sites are shown. These reactions contained 1.7×10^{-10} M of the *SM50* probe containing the tandem P3A consensus sites at positions −128 to −102 (see Table 2), and 0.1 μl of P3A2 purified with an *SM50* column (2 cycles, Table 1). (A) Competition series with the same binding sites, represented by concatenated oligonucleotides 25/26. Reactions were assembled that contained the amount of concatenate indicated below each lane (see Materials and methods for oligonucleotide sequences). The graph in the lower section of the figure shows a determination of K_D and P_0, the concentration of active P3A2 in the reaction, using the relation

$$\frac{PD_s}{D_s} = -\frac{1}{K_D} \times PD_s + \frac{P_0}{K_D} \qquad (1)$$

Here PD_s represents the concentration of factor: site complex, and D_s represents free probe concentration. The value of K_D determined in this experiment is 1.7×10^{-9} M. P_0 was 7.2×10^{-10} M. It was estimated that about 28% of the protein in the reaction was bound to the large excess of nonspecific poly (dAT) competitor present in the reaction (K_D for this nonspecific reaction is expected to be about 1.8×10^{-3} M [Calzone *et al.* 1988], and the poly (dAT) was present at 4.2×10^{-4} M) (B) Competition series with *CyIIIa* binding site at −182 to −195, represented by concatenated oligonucleotide 11/12 as competitor, at the concentrations indicated. Results were analyzed as follows, where the subscript 1 refers to the *SM50* site:P3A2 reaction, and the subscript 2 refers to the *CyIIIa* site:P3A2 reaction:

$$\frac{PD_{s1}}{D_{s1}} = \frac{K_{D2}}{K_{D1}} \times \frac{PD_{s2}}{D_{s2}} \qquad (2)$$

Here PD_{s1}/D_{s1} is given by the ratio of bound to free probe cts min^{-1}; K_{D1} was obtained from eq. (1);

$$PD_{s2} = P_0 - PD_{s1} - P, \quad \text{where} \quad P = \frac{K_{D1} \times PD_{s1}}{D_{s1}}$$

and D_{s2} is the concentration of oligonucleotide 11/12 in the reaction (actually it is this concentration less PD_{s2}, but the correction is insignificant (<1%)). Equations 1 and 2 follow from the definition of K_D. The value of K_{D2}/K_{D1} obtained from the experiment shown was 58.

the *effective protein concentration required for P3A2 function could vary widely for different regulatory DNAs*. Different genes would thus respond differently to given P3A2 concentrations within the same nucleus.

Discussion

In our view the sea urchin embryo presents an exciting opportunity to characterize by direct methods the major *trans*-regulatory molecules that organize the initial diversification of cell function. The lineage of the embryo is known (Cameron *et al.* 1987; Cameron and Davidson, 1991), and a specific pattern of interblastomere inductions that might account for the crucial early founder cell specifications can be inferred (Hörstadius, 1939; Davidson, 1989). A number of territory-specific genes that serve as markers for founder cell specification have been cloned. The molecular interactions that are causally responsible for territorial marker gene regulation could thus provide an image of the mechanism by which the initial territorial specifications take place. However, the usefulness of this approach depends entirely on the feasibility of cloning gene

regulatory factors, given only the target sites that constitute the *cis*-regulatory domains of key marker genes.

In this work we report what is to our knowledge the first application of site-specific affinity chromatography to cloning of regulatory factors from sea urchin embryos. The accompanying paper (Höög *et al.* 1991) describes the application of a direct ligand screening method for obtaining factor cDNA, using the same P3A target sites. Starting with target site sequences, these are the two major routes to regulatory factor characterization, i.e. other than use of homologous probes derived from preexistant clones. As an example of the latter, a heterospecific, homologous probe was used to isolate the sea urchin version of the USF factor, for which there is a target site in the *SpecI* gene (Tomlinson *et al.* 1990). It will be particularly interesting to compare sea urchin regulatory proteins isolated without any *a priori* dependence on homology, to those utilized by mammals, as the echinoderms are the only subchordate deuterostome group for which we possess significant molecular level data. Perhaps not surprisingly, the P3A2 protein described in this paper turned out to belong to no previously known class of regulatory factors. The ligand screening method applied by Höög *et al.* (1991) revealed a second factor that we call P3A1, which recognizes almost the same target sites, and is clearly a member of the Zn finger family of DNA-binding proteins. We discuss the possible interplay of P3A1 and P3A2 in the following paper.

The factor purification procedures and their practical implications

Sea urchin embryos can be raised synchronously in enormous quantities. In this work, we describe the preparation of nuclear extract from 2.5×10^9 embryos, or about 1×10^{12} nuclei. Subsequently we processed about 10× this quantity of embryos. We have found that the embryos can be frozen, and nuclear pellets and soluble protein extract obtained at later convenience. The extract itself is stable when stored frozen, and can be used for at least two years without notable loss of specific DNA-binding activity. P3A2 presented a signficant challenge for purification since it is a very low abundance factor, particularly in the 24 h embryos from which the nuclear extract was obtained. Calzone *et al* (1988) estimated that at its peak in mid-late cleavage the average blastomere nucleus contains about 3000 molecules of P3A2, but there are only about 1200 molecules/nucleus at blastula stage. However, even at this low concentration easily available quantities of nuclear extract suffice Thus 10^9, 24 h embryos would contain about 0.8 nanomole of P3A2, and this is more than adequate for recovery by standard methods of enough sequence to generate cloning probes. Peptide separation and mass spectrometer sequencing instrumentation now available require only 0.1–0.2 nanomoles, and current improvements in technology are rapidly lowering these limits. We have discovered, furthermore, that the same extract can be *sequentially* passed over a series of affinity columns bearing target sites for different factors. Thus a set of diverse factors can be purified from a single aliquot of nuclear extract. Recently we have succeeded in automating sequential affinity chromatography, as will be described elsewhere. The relevant point is that for the sea urchin embryo the quantities of material and the technology and methods already in hand provide ready access to any regulatory factor that in crude nuclear extracts will bind tightly and specifically to its DNA target site.

The purification of P3A2 is instructive from a quantitative point of view. The initial steps are important, since the nuclear pellet retains only 2.5 % of the total embryo protein and the nuclear extract itself includes only about 10 % of total nuclear pellet protein; hence, we estimate that a 400-fold purification has been attained even before affinity chromatography. Table 1 reports about 1600-fold purification in the affinity chromatography step (with about 50 % yield). The combined purification factor for nuclear P3A2 would thus be about 6×10^5. This was probably several-fold more purification than actually would have been required, as noted in text, and in the event we obtained far more sequence than was necessary, as illustrated in Fig. 4.

P3A2: Different affinities for target sites in CyIIIa, Spec*I and* SM50 *genes*

Although we yet lack any coherent understanding of the provenance and distribution of active form(s) of P3A2 in the cleavage-stage embryo, initial observations suggest that several different phenomena may be involved, some or all of which could be functionally important. (*i*) P3A2 exists in several different charge isoforms, as revealed by high resolution, two-dimensional gel electrophoresis (M. Harrington and F. Calzone, unpublished). Thus P3A2 is subject to modification, probably phosphorylation. Such modifications could of course affect biological activity, as predicted to account for the conditional specification of gene expression in the aboral ectoderm founder cells (Davidson, 1989). (*ii*) The same anti-P3A2 antibody used for the experiments of Fig. 6 detects P3A2 in homogenates of unfertilized eggs (R. Zeller, unpublished), though we were unable to demonstrate any maternal DNA-binding activity for P3A2 (Calzone *et al.* 1988). The maternal factor could be sequestered, complexed with an inhibitor, or modified in specific ways. (*iii*) P3A2 is also encoded by a low abundance maternal mRNA (Cutting *et al.* 1990). Whether in later embryos it is of maternal or zygotic origin, the steady state content of this mRNA per embryo remains essentially unchanged throughout early development. Cutting *et al.* (1990) showed that the amount of P3A2 that can be translated on the P3A2 mRNA could account for the total P3A2 measured in late cleavage nuclei and, of importance in this context, that the newly synthesized P3A2 would be present in much higher concentration in macromere than in micromere nuclei if the message were evenly distributed per unit volume of embryo cytoplasm. Thus active P3A2 concentration could be regulated by several different mechanisms in

the early embryo. An obvious implication is that some or all of these mechanisms are used to set up a crucial spatial distribution of active, intranuclear P3A2 concentrations, with respect to particular sets of cleavage-stage founder cells. Here the observations summarized in Table 2 of this paper are directly relevant. The relative equilibrium constants that we obtained of course represent an average for whatever P3A2 variants are present in 24 h embryo nuclei. Nonetheless, these measurements reveal sharp preferences for target sites in some genes relative to others. There are two different aspects to this phenomenon. The closely spaced double sites of both the *SM50* and *Spec*I genes display 30- to 50-fold higher affinity than the single *CyIIIa* site used as a standard in Table 2. This is unlikely to be due to interaction between P3A2 molecules since the sites are spaced differently in these two promoters, and in the *SM50* gene are not even located on the same side of the DNA helix. Perhaps each site acts as a one-dimensional diffusion 'concentrator' for the other. If it depends on one-dimensional diffusion, this effect would be expected to drop off rapidly with intersite distance (Berg and Von Hippel, 1985). However, a second explanation is required as well, *viz* that sequence elements outside of the target site core also affect the equilibrium dissociation behavior of the complexes. Thus Table 2 shows that a high relative equilibrium constant is not observed for concatenated oligonucleotide 11/12, representing the standard *CyIIIa* site, though in these concatenates there are again closely spaced target sites, here only 25 bp apart. Furthermore, the *single* proximal *SM50* target site also displays a high relative equilibrium constant, and with respect to the others listed in Table 2 this site is distinguished only by its external sequence features.

Whether the results of Table 2 are due to target site organization or to target site sequence, these features of the *cis*-regulatory DNA of different genes would give rise to different *functional interpretations of limiting intranuclear P3A2 concentrations*. If given concentrations of P3A2 are indeed specific to given sets of lineage founder cells, these different interpretations could determine which genes will be subject to P3A2 regulation in which blastomeres

We would like to acknowledge the assistance of Patrick Leahy in the culture of the embryos. This research was supported by an NIH Grant [HD-05753] to E.H.D.; by an NSF Grant [DCB-8912530], an ACS Institutional Grant [INS-166B], and Funds from the Academic Committee on Research at UCI, to F.J.C.; and by an NSF Grant [DIR-8618937] to D.B.T. C.H was supported by a postdoctoral fellowship from EMBO, and R.W Z is supported by an NIH Training Grant [GM-07616].

References

AEBERSOLD, R H., NIKA, H., PIPES, G D , WETTENHALL, R E. H , CLARK, S. M., HOOD, L. E AND KENT, S B H. (1989) Accelerated high sensitivity sequence analysis of proteins and peptides immobilized on chemically-modified glass fiber discs. In *Methods in Protein Sequence Analysis* (ed. B Wittmann-Liebold), pp 79–97. Springer-Verlag

BENSON, S C , SUCOV, H M., STEPHENS, L , DAVIDSON, E H AND WILT, F (1987) A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule I Authentication of the cloned gene and its developmental expression *Devl Biol* **120**, 499–506

BERG, O. G AND VON HIPPEL, P H (1985) Diffusion-controlled macromolecular interactions *Ann Rev Biophys Biophys Chem* **14**, 131–160

BRADFORD, M (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding *Anal Biochem* **72**, 248–254.

CALZONE, F J , THÉZÉ, N , THIEBAUD, P , HILL, R L , BRITTEN, R J AND DAVIDSON, E H (1988) Developmental appearance of factors that bind specifically to *cis*-regulatory sequences of a gene expressed in the sea urchin embryo. *Genes Dev* **2**, 1074–1088

CAMERON, R A , BRITTEN, R J AND DAVIDSON, E H (1989) Expression of two actin genes during larval development in the sea urchin *Strongylocentrotus purpuratus Mol Reprod and Dev* **1**, 149–155

CAMERON, R. A AND DAVIDSON, E H (1991) Cell type specification during sea urchin development *Trends in Genetics* (in press)

COX, K H , ANGERER, L M , LEE, J J , BRITTEN, R J , DAVIDSON, E H AND ANGERER, R C (1986) Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis *J molec Biol* **188**, 159–172

CUTTING, A E , HÖÖG, C , CALZONE, F J , BRITTEN, R J AND DAVIDSON, E H (1990). Rare maternal mRNAs code for regulatory proteins that control lineage specific gene expression in the sea urchin embryo *Proc natn Acad Sci U S A* **87**, 7953–7959.

DAVIDSON, E. H (1986) *Gene Activity in Early Development* Academic Press· Orlando, Florida

DAVIDSON, E. H (1989) Lineage-specific gene expression and the regulative capacities of the sea urchin embryo· A proposed mechanism *Development* **105**, 421–445

DENHARDT, D T (1966) A membrane-filter technique for the detection of complementary DNA *Biochem biophys Res Commun* **23**, 641–646

FLYTZANIS, C N , BRITTEN, R J AND DAVIDSON, E H (1987) Ontogenic activation of a fusion gene introduced into the sea urchin egg. *Proc natn Acad Sci U S A* **84**, 151–155

GAN, L , ZHANG, W. AND KLEIN, W H (1990) Repetitive DNA sequences linked to the sea urchin Spec genes contain transcriptional enhancer-like elements *Devl Biol* **139**, 186–196

HARDIN, S H , CARPENTER, C D , HARDIN, P E , BRUSKIN, A M AND KLEIN, W H (1985) Structure of the *Spec1* gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus J molec Biol* **186**, 243–255

HÖÖG, C , CALZONE, F C , CUTTING, A E , BRITTEN, R J AND DAVIDSON, E H. Gene regulatory factors of the sea urchin II. Two dissimilar proteins, P3A1 and P3A2, bind to the same target sites that are required for early territorial gene expression *Development* **112**, 000–000

HÖRSTADIUS, S (1939) The mechanics of sea urchin development, studied by operative methods *Biol Rev Cambridge Phil Soc* **14**, 132–179

HOUGH-EVANS, B. R , BRITTEN, R J AND DAVIDSON, E H. (1988) Mosaic incorporation and regulated expression of an exogenous gene in the sea urchin embryo *Devl Biol* **129**, 198–208

HOUGH-EVANS, B R , FRANKS, R R , CAMERON, R. A , BRITTEN, R. J AND DAVIDSON, E H (1987). Correct cell type-specific expression of a fusion gene injected into sea urchin eggs *Devl Biol* **121**, 576–579

HOUGH-EVANS, B R , FRANKS, R R , ZELLER, R W., BRITTEN, R J AND DAVIDSON, E H (1990). Negative spatial regulation of the lineage specific *CyIIIa* actin gene in the sea urchin embryo. *Development* **108**, 41–50

IKEGAKI, N AND KENNET, R. H (1989). Glutaraldehyde fixation of the primary antibody-antigen complex on nitrocellulose paper

increases the overall sensitivity of immunoblot assay. *J Imm. Meth* **124**, 205–210

KADONAGA, J T., CARNER, K R , MASIARZ, F R. AND TJIAN, R (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain *Cell* **51**, 1079–1090

LAEMMLI, U K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4 *Nature* **227**, 680–685

LYNN, D , ANGERER, L , BRUSKIN, A , KLEIN, W AND ANGERER, R (1983) Localization of a family of mRNAs in a single cell type and its precursors in sea urchin embryos *Proc natn Acad. Sci U S A.* **80**, 2656–2660.

PARKER, C. S AND TOPOL, J (1984) A *Drosophila* RNA polymeraseII transcription factor binds to the regulatory site on an hsp70 gene *Cell* **37**, 273–283

PIRROTTA, V , MANET, E , HARDON, E , BICKEL, S. E. AND BENSON, M (1987) Structure and sequence of the *Drosophila* zeste gene *EMBO J* **6**, 791–799

STUDIER, F W AND MOFFAT, B A (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes *J molec Biol* **189**, 113–130

STURM, R A , DAS, G AND HERR, W (1988) The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain *Genes Dev* **2**, 1582–1599.

SUCOV, H M , BENSON, S , ROBINSON, J. J , BRITTEN, R. J , WILT, F AND DAVIDSON, E H. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule II Structure of the gene and derived sequence of the protein *Devl Biol* **120**, 507–519

SUCOV, H , HOUGH-EVANS, B R , FRANKS, R. R , BRITTEN, R J AND DAVIDSON, E H (1988). A regulatory domain that directs lineage specific expression of a skeletal matrix protein gene in the sea urchin embryo *Genes Dev* **2**, 1238–1250

THÉZÉ, N , CALZONE, F J , THIEBAUD, P , HILL, R L , BRITTEN, R J AND DAVIDSON, E H (1990) Sequence of the *CyIIIa* actin gene regulatory domain bound specifically by sea urchin embryo nuclear protein. *Mol Reprod and Dev* **25**, 110–122

THIEBAUD, P , GOODSTEIN, M , CALZONE, F J , THÉZÉ, N , BRITTEN, R J AND DAVIDSON, E H (1990) Intersecting batteries of differentially expressed genes in the sea urchin embryo *Genes Dev* **4**, 1999–2010

TOMLINSON, C. R., KOZLOWSKI, M. T AND KLEIN, W. H. (1990). Ectoderm nuclei from sea urchin embryos contain a Spec-DNA binding protein similar to the vertebrate transcription factor USF *Development* **110**, 259–272

VINSON, C R , LAMARCO, K L., JOHNSON, P F , LANDSCHULTZ, W H AND MCKNIGHT, S L (1988) *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage *Genes Dev* **2**, 802–806

(Accepted 6 February 1991)